CONTENT IDENTIFIERS TRIGGERING CORRESPONDING RESPONSES THROUGH COLLABORATIVE PROCESSING

Related Application Data

Sub a1

[0001] This application is a continuation-in-part of application 09/858,189, filed May 14, 2001, which is a continuation in part of application 09/571,422, which claimed priority from applications 09/314,648, 09/342,688, 09/342,689, 09/342,971, 09/343,101, 09/343,104, 60/141,468, 60/151,586, 60/158,015, 60/163,332, 60/164,619, 09/531,076, 09/543,125, 09/547,664, and 09/552,998.

[0002] This application is also a continuation-in-part of copending applications 09/574,726 and 09/476,686, both of which claim priority to application 60/134,782.

[0003] The present application claims priority benefit to the foregoing applications.

[0004] The subject matter of this application is also related to that of 09/620,019, 60/257,822, 60/232,163, and 09/404,291.

Field of the Invention

[0005] The present invention relates to computer-based systems, and more particularly relates to systems that identify electronic or physical objects (e.g., audio, printed documents, video, etc.), and trigger corresponding responses.

Background

[0006] In application 09/571,422 (now laid-open as PCT publication WO 00/70585), the present assignee described technology that can sense an object identifier from a physical or electronic object, and trigger a corresponding computer response.

[0007] In applications 09/574,726 and 09/476,686, the present assignee described technology that uses a microphone to sense audio sounds, determine an identifier corresponding to the audio, and then trigger a corresponding response.

Detailed Description

[0008] Although the cited patent applications focused on use of digital watermarks to identify the subject objects/audio, they noted that the same applications and benefits can be provided with other identification technologies.

[0009] One such suitable technology - variously known as robust hashing, fingerprinting, etc. – involves generating an identifier from attributes of the content. This identifier can then be looked-up in a database (or other data structure) to determine the song (or other audio track) to which it corresponds.

[0010] Various fingerprinting technologies are known. For example, a software program called TRM, from Relatable Software, was written up in the Washington Post as follows:

> *TRM performs a small technological miracle: It "fingerprints" songs, analyzing beat and tempo to generate a unique digital identifier. Since every song is slightly different, no two "acoustic fingerprints" are alike, not even live and studio versions of the same melody.*

[0011] Tuneprint is another such audio fingerprinting tool. Tuneprint is understood to utilize a model of human hearing used to predict how audio will appear after it's been distorted by the human ear, and the parts of neural processing that are understood. This is some of the same information that led to MP3 encoders achieving exceptional audio compression. Characteristics that uniquely identify the track are then identified by picking out the most important, surprising, or significant features of the sound.

[0012] Yet another fingerprinting program is Songprint, available as an open source library from freetantrum.org.

[0013] Still other fingerprinting technologies are available from Cantametrix (see, e.g., published patent applications WO01/20483 and WO01/20609).

[0014] One particular approach to fingerprinting is detailed in the present assignee's application 60/263,490, filed January 22, 2001.

[0015] One form of fingerprint may be derived by applying content – in whole or part, and represented in time- or frequency format – to a neural network, such as a Kohonen self-organizing map. For example, a song may be identified by feeding the first 30 seconds of audio, with 20 millisecond Fourier transformed windows, into a Kohonen network having 64 outputs. The 64 outputs can, themselves, form the fingerprint, or they can be further processed to yield the fingerprint.

[0016] A variety of other fingerprinting tools and techniques are known to artisans in the field. Others are disclosed, e.g., in applications 60/257,822, 09/563,664, and 09/578,551. See also the chapter on Fingerprinting by John Hyeon Lee, in Information Hiding: Techniques for Steganography and Digital Watermarking edited by Stefan Katzenbeisse and Fabien A.P. Petitcolas, published by Artech House.

[0017] One way to generate a fingerprint is to "hash" the audio, to derive a shorter code that is dependent, in a predetermined way, on the audio data. However, slight differences in the audio data (such as sampling rate) can cause two versions of the same song to yield two different hash codes. While this outcome is advantageous in certain outcomes, it is disadvantageous in many others.

[0018] Generally preferable are audio fingerprinting techniques that yield the same fingerprints, even if the audio data are slightly different. Thus, a song sampled at a 96K bit rate desirably should yield the same fingerprint as the same song sampled at 128K. Likewise, a song embedded with steganographic watermark data should generally yield the same fingerprint as the same song without embedded watermark data.

[0019] One way to do this is to employ a hash function that is insensitive to certain changes in the input data. Thus, two audio tracks that are acoustically similar will hash to the same code, notwithstanding the fact that individual bits are different. A variety of such hashing techniques are known.

[0020] Another approach does not rely on "hashing" of the audio data bits. Instead, the audio is decomposed into elements having greater or lesser perceptibility. Audio compression techniques employ such decomposition methods, and discard the elements that are essentially imperceptible. In fingerprinting, these elements can also be disregarded, and the "fingerprint" taken from the acoustically significant portions of the audio (e.g., the most significant coefficients after transformation of the audio into a transform domain, such as DCT).

[0021] Some fingerprinting techniques do not rely on the absolute audio data (or transformed data) per se, but rather rely on the changes in such data from sample to sample (or coefficient to coefficient) as an identifying hallmark of the audio.

[0022] Some fingerprinting algorithms consider the entire audio track (e.g., 3 minutes). Others work on much shorter windows – a few seconds, or fractions of seconds. The former technique yields a single fingerprint for the track. The latter yields plural fingerprints – one from each excerpt. (The latter fingerprints can be concatenated, or otherwise combined, to yield a master fingerprint for the entire audio track.) For compressed audio, one convenient unit from which excerpts can be formed is the frame

or window used in the compression algorithm (e.g., the excerpt can be one frame, five frames, etc.).

[0023] One advantage to the excerpt-based techniques is that a song can be correctly identified even if it is truncated. Moreover, the technique is well suited for use with streaming media (in which the entire song data is typically not available all at once as a single file).

[0024] In database look-up systems employing fingerprints from short excerpts, a first fingerprint may be found to match 10 songs. To resolve this ambiguity, subsequent excerpt-fingerprints can be checked.

[0025] One way of making fingerprints "robust" against variations among similar tracks is to employ probabilistic methods using excerpt-based fingerprints. Consider the following, over-simplified, example:

Fingerprinted excerpt	Matches these songs in database
Fingerprint 1	A, B, C
Fingerprint 2	C, D, E
Fingerprint 3	B, D, F
Fingerprint 4	B, F, G

[0026] This yields a "vote" tally as follows:

Matches to	A	B	C	D	E	F	G
# Hits	1	3	2	2	1	2	1

[0027] In this situation, it appears most probable that the fingerprints correspond to song B, since three of the four excerpt-fingerprints support such a conclusion. (Note that one of the excerpts - that which yielded Fingerprint 2 - does not match song B at all.)

[0028] More sophisticated probabilistic techniques, of course, can be used.

[0029] Once a song has been identified in a database, a number of different responses can be triggered. One is to impose a set of usage controls corresponding to terms set by the copyright holder (e.g., play control limitations, record control, fee charges, etc.) Another is to identify metadata related to the song, and provide the metadata to a user (or a link to the metadata). In some such applications, the song is simply identified by title and artist, and this information is returned to the user, e.g., by email, instant messaging, etc. With this information, the user can be given an option to purchase the music in CD or electronic form, purchase related materials (t-shirts, concert tickets), etc. A great variety of other content-triggered actions are disclosed in the cited applications.

[0030] One of the advantages of fingerprint-based content identification systems is that they do not require any alteration to the content. Thus, recordings made 50 years ago can be fingerprinted, and identified through such techniques.

[0031] Going forward, there are various advantages to encoding the content with the fingerprint. Thus, for example, a fingerprint identifier derived from a song can be stored in a file header of a file containing that song. (MP3 files, MPEG files, and most other common content file formats include header fields in which such information can readily be stored.) The fingerprint can then obtained in two different ways – by reading the header info, and by computation from the audio information. This redundancy offers several advantages. One aids security. If a file has a header-stored fingerprint that does not match a fingerprint derived from the file contents, something is amiss – the file may be destructive (e.g., a bomb or virus), or the file structure may misidentify the file contents.

[0032] In some embodiments, the fingerprint data (or watermark data) stored in the header may be encrypted, and/or authenticated by a digital signature such as a complete hash, or a few check bits or CRC bits. In such cases, the header data can be the primary source of the fingerprint (watermark) information, with the file contents being processed to re-derive the fingerprint (watermark) only if authentication of the fingerprint stored in the header fails. Instead of including the fingerprint in the header, the header can include an electronic address or pointer data indicating another location (e.g., a URL or database record) at which the fingerprint data is stored. Again, this information may be secured using known techniques.

[0033] Similarly, the fingerprint can point to a database that contains one or more IDs that are added via a watermark. This is useful when CDs are being converted to MP3 files (i.e. ripped) and the fingerprint is calculated from a hash of the table of contents (TOC) such as done with CDDB.com, or from all of the songs. In this case, the database entry for that fingerprint could include a list of IDs for each song, and these IDs are added via a watermark and/or frame header data. This can also be useful where the content is identified based upon a group of fingerprints from plural excerpts, in which case the database that determines the content also contains an identifier, unrelated to the fingerprint(s) for that piece of content that can be embedded via a watermark.

[0034] Instead of, or in addition to, storing a fingerprint in a file header, the fingerprint data may be steganographically encoded into the file contents itself, using known watermarking techniques (e.g., those disclosed in application 09/503,881, and patents 6,061,793, 6,005,501 and 5,940,135). For example, the fingerprint ID can be duplicated in the data embedded via a watermark.

[0035] In some arrangements, a watermark can convey a fingerprint, and auxiliary data as well. The file header can also convey the fingerprint, and the auxiliary data. And even

if the file contents are separated from the header, and the watermark is corrupted or otherwise lost, the fingerprint can still be recovered from the content. In some cases, the lost auxiliary data can alternatively be obtained from information in a database record identified by the fingerprint (e.g., the auxiliary information can be literally stored in the record, or the record can point to another source where the information is stored).

[0036] Instead of especially processing a content file for the purpose of encoding fingerprint data, this action can be done automatically each time certain applications process the content for other purposes. For example, a rendering application (such as an MP3 player or MPEG viewer), a compression program, an operating system file management program, or other-purposed software, can calculate the fingerprint from the content, and encode the content with that information (e.g., using header data, or digital watermarking). It does this while the file is being processed for another purpose, e.g., taking advantage of the file's copying into a processing system's RAM memory, from slower storage.

[0037] In formats in which content is segregated into portions, such as MP3 frames, a fingerprint can be calculated for, and encoded in association with, each portion. Such fingerprints can later be crosschecked against fingerprint data calculated from the content information, e.g., to confirm delivery of paid-for content. Such fingerprints may be encrypted and locked to the content, as contemplated in application 09/620,019.

[0038] In addition, in this frame based systems, the fingerprint data and/or watermark data can be embedded with some or all data throughout each frames. This way a streaming system can use the header to first check the song for identification, and if that identification is absent or not authenticated, the system can check for the watermark and/or calculate the fingerprint. This improves the efficiency and cost of the detecting system.

[0039] Before being encrypted and digitally signed, the data in the frame header can be modified by the content, possibly a hash of the content or a few critical bits of content. Thus, the frame header data cannot be transferred between content. When reading the data, it must be modified by the inverse transform of the earlier modification. This system can be applied whether the data is embedded throughout each frame or all in a global file header and is discussed in application serial 09/404,291 entitled "Method And Apparatus For Robust Embedded Data" by Ken Levy on 9/23/99. Reading this secure header data is only slightly more complex than without the modification, such that the system is more efficient than always having to calculate the fingerprint and/or detect the watermark.

Collaboration

[0040] In some situations, content may be processed by plural users, at about the same time, to generate corresponding identifiers. This may occur, for example, where the content is a song or advertisement broadcast over the radio. Many listeners in a metropolitan area may process audio from the same song broadcast over the radio, e.g., to learn the artist or song title, to engage in some related e-commerce activity, or for another purpose (such as the other purposes identified in the cited applications).

[0041] In such cases it may be desirable to employ collaboration between such users, e.g., to assure more accurate results, to reduce the processing burden, etc.

[0042] In one embodiment, each user generates several different fingerprints from the content (such as those identified in the table, above). These fingerprints may be aggregated with other fingerprints submitted from other users within a given time window (e.g., within the past twenty seconds, or within the past fifteen and next five seconds). Since more data is being considered, the "correct" match may more likely stand out from spurious, incorrect matches.

[0043] Consider Users 1 and 2, whose content yields fingerprints giving the following matches (User 1 is unchanged from the earlier example):

Fingerprinted excerpt	Matches these songs in database
User 1, Fingerprint N	A, B, C
User 1, Fingerprint N+1	C, D, E
User 1, Fingerprint N+2	B, D, F
User 1, Fingerprint N+3	B, F, G
User 2, Fingerprint M	A, B, E
User 2, Fingerprint M+1	H, I, A
User 2, Fingerprint M+2	X, Y, Z

[0044] Aggregating the fingerprints from the two users results in an enhanced vote tally in which song B is the evident correct choice – with a higher probability of certainty than in the example earlier given involving a single user:

Matches to	A	B	C	D	E	F	G	H	I	X	Y	Z
# Hits	2	4	2	2	2	2	1	1	1	1	1	1

[0045] Moreover, note that User 2's results are wholly ambiguous – no song received more than a single candidate match. Only when augmented by consideration of fingerprints from User 1 can a determination for User 2 be made.

This collaboration aids the situation where several users are listening to the same content. If two users are listening to different content, it is highly probable that the fingerprints of the two users will be uncorrelated. No benefit arises in this situation, but the collaboration does not work an impairment, either. (In identifying the song for User 1, the system would only check the candidates for whom User 1 voted. Thus, if the above table showed 5 votes for a song J, that large vote count would not be considered in

identifying the song for User 1, since none of the fingerprints from User 1 corresponded to that song.)

[0046] It will be recognized that the different fingerprints obtained by different users from the same song may be due to a myriad of different factors, such as ambient noise, radio multipath reception, different start times for audio capture, etc.

[0047] In the example just given, the number of fingerprints computed for each user can be reduced when compared with non-collaborative approaches, while still providing enhanced confidence in the final song determination.

[0048] Another collaborative embodiment employs a reference system. Consider again the example of radio broadcasts in a metropolitan area. Reference receivers can be installed that continuously receive audio from each of several different radio stations. Instead of relying on sound picked up by a microphone from an ambient setting, the reference receivers can generate fingerprint data from the audio in electronic form (e.g., the fingerprint-generation system can be wired to the audio output of the receiver). Without the distortion inherent in rendering through a loudspeaker, sensing through a microphone, and ambient noise effects, more accurate fingerprints may be obtained.

[0049] The reference fingerprints can be applied to the database to identify – in essentially real-time and with a high degree of certainty - the songs (or other audio signals) being broadcast by each station. The database can include a set of fingerprints associated with the song. Alternatively, the reference receiver can generate fingerprints corresponding to the identified song.

[0050] Consumers listen to audio, and fingerprints are generated therefrom, as before. However, instead of applying the consumer-audio fingerprints to the database (which may involve matching to one of hundreds of thousands of possible songs), the consumer fingerprints are instead compared to the fingerprints generated by the reference receivers

(or songs determined there from). The number of such reference fingerprints will be relatively low, related to the number of broadcast stations being monitored. If a consumer-audio fingerprint correlates well with one of the reference fingerprints, then the song corresponding to that reference fingerprint is identified as the song to which the consumer is listening. If the consumer-audio fingerprint does not correlate well with any of the reference fingerprints, then the system may determine that the audio heard by the consumer is not in the subset monitored by the reference receivers, and the consumer-audio fingerprints can thereafter be processed against the full fingerprint database, as earlier described.

[0051] The system just described is well suited for applications in which the geographical location of the consumer is known, or can be inferred. For example, if the consumer device that is listening to the audio is a cell phone, and the cellular wireless infrastructure is used to relay data with the phone, the cell system can determine whether the geographical location of the listener (e.g., by area code, cell site, etc.). (Use of such cell-system data to help geographically locate the user can be employed advantageously in several such song-identification systems.).

[0052] Even if the consumer's location cannot be determined, the number of songs playing on radio stations nationwide is still a small subset of the total number of possible songs. So a nationwide system, with monitoring stations in many metropolitan areas, can be used to advantage.

[0053] As an optional enhancement to such a collaborative system, broadcast signals (e.g., audio signals) are digitally watermarked. The digital watermark preferably contains plural-bit data, which is used to identify the audio signal (e.g., a set of audio fingerprints from the audio signal, song title, copyright, album, artist, and/or record label, etc., etc.). The plural-bit data can either directly or indirectly identify the audio signal. In the indirect case, the plural-bit data includes a unique identifier, which can be used to interrogate a database. The database preferably includes some or all of the identifying

information mentioned above. A reference receiver decodes an embedded digital watermark from a received audio signal. The unique identifier is used to interrogate the database to identify a fingerprint or a set of fingerprints associated with the particular audio signal. In some cases, the set includes one fingerprint; in other cases, the set includes a plurality of fingerprints. On the user side, fingerprints are generated and relayed to the reference receiver (or associated interface). The user's fingerprints are then compared against the reference fingerprints, as discussed above in the earlier embodiments.

[0054] The foregoing are just exemplary implementations of the present invention. It will be recognized that there are a great number of variations on these basic themes. The foregoing illustrates but a few applications of the detailed technology. There are many others.

[0055] To provide a comprehensive disclosure without unduly lengthening this specification, applicants incorporate by reference the patents and patent applications cited above. It is applicant's express intention to teach that the methods detailed herein are applicable in connection with the technologies and applications detailed in these cited patents and applications.

[0056] Although the foregoing specification has focused on audio applications, it will be recognized that the same principles are likewise applicable with other forms of content, including still imagery, motion pictures, video, etc. References to "songs" are illustrative only, and are not intended to limit the present invention. The inventive methods and systems could also be applied other audio, image, video signals as well. Also, for example, Digimarc MediaBridge linking from objects to corresponding internet resources can be based on identifiers derived from captured image data or the like, rather than from embedded watermarks. As such, the technique is applicable to images and video.